Facial Response to Video Content in Depression

Gordon McIntyre
Faculty of Information Sciences and Engineering
University of Canberra, Australia
gordon.mcintyre@canberra.edu.au

Roland Goecke
Faculty of Information Sciences and Engineering
University of Canberra, Australia
roland.goecke@ieee.org

Michael Breakspear
School of Psychiatry
University of New South Wales, Australia
Black Dog Institute, Sydney, Australia
Division of Mental Health Research
Queensland Institute of Medical Research
Royal Brisbane & Woman's Hospital, Australia
mbreak@unsw.edu.au

Gordon Parker
School of Psychiatry
University of New South Wales, Australia
Black Dog Institute, Sydney, Australia
g.parker@unsw.edu.au

ABSTRACT

Depressed subjects have been shown to respond differently to images of positive and negative content, when compared with non-depressed subjects. The underlying cause could be the impaired inhibition of negative affect, which has been found in depressed patients across several studies. We describe the techniques used in an ongoing study to compare the clinical diagnosis of depression with automatically measured facial activity and expressions. Video recordings are made of patients and control subjects watching a series of film clips, portraying negative and positive content. Subject-specific Active Appearance Models are built in order to extract visual features from the faces within frames captured from the videos. The raw feature data is then used to measure each participant‘s facial activity and to train Support Vector Machines for recognition of facial expressions.

1. INTRODUCTION

We are interested in developing an objective measure as an aid for the detection and monitoring of major depressive disorders based on an analysis of facial activity. In this preliminary study, 13 participants, 6 controls and 7 patients, were recorded using a high-quality video camera, while viewing affective content and answering emotive questions. The movies clips together with intended induced emotion, rated for their affective content by [2], are listed in Table 1. Table 2 summarises the diagnosis of each subject in the study.

Section 2 describes the methodology, followed by an outline of the techniques used to measure facial activity and expressions in Sections 3 and 4, resp. Section 5 concludes the paper with a short discussion of the results and ongoing work.

Movie	(Emotion)	Length (mm:ss)
Bill Cosby	(Happy)	02:06
The Champ	(Sad)	02:49
Weather	(Happy)	00:57
Silence of the Lambs	(Fear)	03:44
Cry Freedom	(Anger)	02:40
The Shining	(Fear)	01:07
Capricorn One	(Surprise)	00:49

Table 1: Paradigm movie clip list

2. METHODOLOGY

The experimental affect induction paradigm [4] is presented to participants via an interactive computer package [5]. Videos of the subjects viewing the paradigm are recorded with an Allied Vision Technologies Pike 100C camera. The camera is connected to an Apple MacBook Pro.

Once the video has been recorded, selected frames depicting facial variations are captured,[1] and used to 1) build a person-specific Active Appearance Model (AAM, [6]) for the participant;[2] and 2) construct a Support Vector Machine (SVM) classifier for each participant's emotional expressions.

The AAMs and SVMs are then applied to frames, captured from the video at 200 ms intervals (this seemed a reasonable choice of interval based, anecdotally, on the speed of movement of facial features, however, there is no evidence that this is the optimal rate). As each frame is captured, the images are scanned using the Viola and Jones [7] technique to detect the global location of the face within the image. Within the global coordinates, the AAM is used to track and measure the facial shape and texture features. "Shape" refers to the set of landmark points, which are captured as a set of normalised (x, y) Cartesian coordinates. The AAM output features are then used to classify facial expressions using an SVM classifier [1]. All outputs are retained within the system to allow for post-processing, e.g. comparisons between participants.

[1] these are selected subjectively at this point in time
[2] person-specific AAMs give better fitting quality [3] and there is no need at this juncture to have generic AAMs

Subject Id	Age	Gender	Diagnosis Clinical	Diagnosis MINI
Co_m_11	39	m		
Co_f_12	26	f		
Co_m_13	33	m		
Co_m_14	31	m		
Co_f_15	31	f		
Co_m_16	20	m		
Pa_m_UP-Mel_05	26	m	UP-MEL	UP-MEL
Pa_f_UP-NonMel_06	34	f	BP2, MEL	UP NON-MEL
Pa_m_Unkown_07	45	m	Unknown	Unknown
Pa_f_UP_BP2_08	32	f	BP2	BP2
Pa_f_UP-NonMel_09	27	f	Unknown	UP-NON MEL
Pa_f_UP-Mel_10	50	f	UP-MEL	UP-MEL
Pa_m_PD_11	53	m	PD	UP-MEL
Total Participants	**6 Co / 7 Pa**			

Table 2: Participant details. Subject IDs in the table take the form XX_G_CD_ID where XX - "Co" for Control or "Pa" for Patient, G - Gender, CD - Diagnosis (Patients only), ID - Sequential ID number (control and patients numbered separately). UP - uni-polar, BP - bi-polar, MEL - melancholic.

3. MEASURING FACIAL ACTIVITY

With the raw feature data captured, an algorithm is used to measure the collective movement of the landmark points between all frames. Although not shown in the algorithm, extreme movements are ignored if they fall outside of predefined thresholds. This is to cater for situations where the face detection in a frame has failed and the AAM "fitting" has not converged, which typically leaves the landmark points scattered around the image.

Algorithm 1: Measuring facial activity

input : set of facial landmark points for all images in video
int i = 0
for *each set of facial landmark points* **do**
  tempx ← distance between x coordinates of current and previous frame
  tempy ← distance between y coordinates of current and previous frame
  `// L1norm is the sum of absolute values`
  FacialActivity[i] ← L1norm(tempx) + one.norm(tempy)
  i++;
output : set of scalar values representing distance between each set of landmark points in video

4. TRACKING FACIAL EXPRESSIONS

The expression, classified by the SVM, is stored within the system for each captured image. To verify that the AAM has fitted properly and that the SVM expression recognition has worked successfully, images marked up with the automatically fitted landmark points can be assembled as an image sequence and played as a short video, as shown in Figure 1.

5. RESULTS AND DISCUSSION

The results to date have been useful in validating the paradigm and technique. To illustrate, Figure 2 shows the facial activity elicited from the Capricorn One movie clip, which is used to induce surprise. Although the figure reveals that patient Pa_f_UP_BP2_08 had a very high degree of facial activity, examination of the video revealed that she displayed non-purposeful or habitual mouth movements throughout the recording. This highlights the need for caution and careful validation of the results.

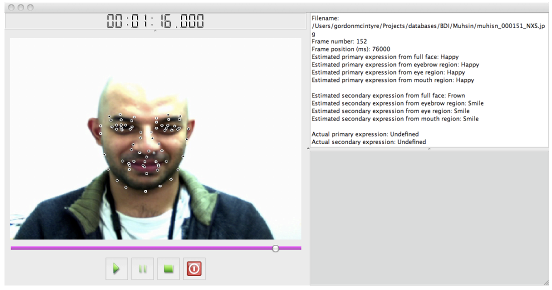

Figure 1: Replaying captured images

As mentioned in the introduction, part of the interview involves participants answering emotive questions. Research is also under way to analyse the participants' vocal response in the question and answer segment, so as to develop a truly multimodal approach. While the results to date are promising, data from more subjects are required before any conclusions can be drawn as to the accuracy of the approach.

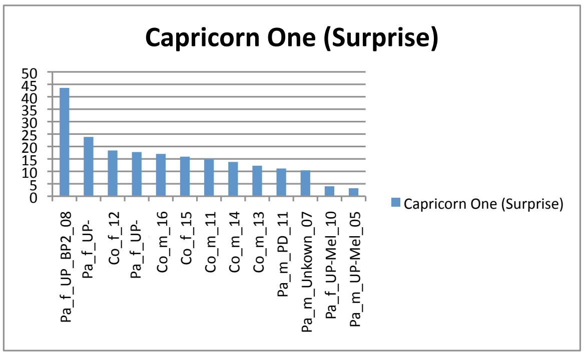

Figure 2: Facial activity response to Capricorn One

6. REFERENCES

[1] C. Chang and C. Lin. LIBSVM: a library for support vector machines. 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[2] J. Gross and R. Levenson. Emotion elicitation using films. *Cognition & Emotion*, 9:87–108, 1995.

[3] R. Gross, I. Matthews, and S. Baker. Generic vs. person specific active appearance models. *Image Vision Computing*, 23(12):1080–1093, 2005.

[4] G. McIntyre, R. Goecke, M. Hyett, M. Green, and M. Breakspear. An approach for automatically measuring facial activity in depressed subjects. In *Proc. ACII2009*, pages 598–605, Sept. 2009.

[5] NBS. Neurobehavioral Systems. `http://www.neurobs.com/`. last accessed 14 April 2010.

[6] J. Saragih and R. Göcke. Learning AAM fitting through simulation. *Pattern Recognition*, 42(11):2628–2636, 2009.

[7] P. Viola and M. Jones. Robust real-time face detection. *Proceedings Eighth IEEE International Conference on Computer Vision ICCV 2001*, 2:747, 2001.